DOCUMENT RESUME

ED 128 425 TM 005 610

AUTHOR Stufflebeam, Daniel L.
TITLE Educational Evaluation: Some Questions and Answers.
PUB DATE [Feb 76]
NOTE 17p.; Paper presented at the Annual Meeting of the American Association of School Administrators (Atlantic City, New Jersey, February 1976)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS *Administrator Guides; Definitions; *Evaluation; Models

ABSTRACT

A conceptual framework that educational administrators may want to adopt as a general guide for their evaluative activities is described. In conceptualizing evaluation, the author sets forth ten questions that anyone trying to understand evaluation might ask and answers them with these statements: (1) evaluation is the ascertainment of merit; (2) evaluation may be applied to a variety of objects; (3) evaluation serves multiple audiences; (4) evaluation serves both decision making and accountability; (5) evaluation assesses goals, plans, activities, and results; (6) evaluation provides both descriptive and judgmental information; (7) evaluation is conducted by both insiders and outsiders; (8) evaluation follows a process of delineating, obtaining, and applying information from the evaluation; (9) evaluation should itself be evaluated for its technical adequacy, utility, ethical treatment of people, and cost effectiveness; and (10 evaluation is defined as the process of delineating, obtaining, and applying descriptive and judgmental information concerning some object's merit as revealed by its goals, plans, process, and product for the purposes of decision making and accountability. (RC)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

TM

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Educational Evaluation:

Some Questions and Answers

An Invited Address at the 1976 National Convention
of the American Association of School Administrators

Atlantic City, New Jersey
February, 1976

by

Daniel L. Stufflebeam
Professor of Educational Leadership
and
Director of the Evaluation Center
College of Education
Western Michigan University

ED128425

TM005 610

Introduction

Educational administrators perform many evaluation tasks. For example, they assess the qualifications and performance of their staff members, the merit of funding requests, and the educational needs and performance of their students. Administrators must implement these and other evaluation responsibilities, both to assure the effectiveness of their school's programs and to be accountable to their constituents.

To perform their evaluation responsibilities, administrators need to be guided by a sound conceptualization of evaluation. In the time alloted to me in this session I will propose and describe a particular conceptual framework.

This framework reflects about a dozen years of evaluation work. This work usually was conducted jointly with educational administrators, and it was conducted with a wide variety of programs at local, state and national levels.

If I have learned anything from my evaluation experiences, it is that educational administrators more often than not are the most important persons in evaluation work. They can stimulate or stifle its implementation. They can provide or deny the conditions that are essential to conduct good evaluation. And they can use, misuse, or ignore evaluation results. Hence, I appreciate this opportunity to propose an evaluation model to the group that I consider most crucial in the performance of quality evaluation.

A Proposed Conceptualization of Evaluation

In conceptualizing evaluation I have found it useful to address ten questions that anyone trying to understand evaluation might ask. These questions are:

1) What is the essential meaning of evaluation?

2) What are the various objects of evaluation work?

3) What audiences should it serve?

4) What are the uses of evaluation findings?

5) What questions are addressed by evaluation?

6) What information does it require?

7) Who should do it?

8) How should they do it?

9) By what standards should their work be judged?

10) What operational definition that reflects the responses to the other questions should guide the evaluation work?

Different evaluators would give different answers to these questions. Moreover, members of this audience may well decide on answers that differ from mine. It is not necessary that every school district adopt the same set of answers to these questions. It is important, however, that each district adopt some set of defensible guidelines and that they put their view of evaluation into operation.

Here again, are the ten questions along with my response to each one:

1) What is the essential meaning of evaluation?

Many definitions have been contrived to respond to this question. Most prominent are those that equate evaluation to measurement, to

experimental research, to relating objectives to outcomes, to servicing decisions, or to rendering professional judgments. As I have argued elsewhere[1] these are inadequate--either because they are too narrow in the questions they address or in the methods they use. What is needed is a basic definition of evaluation that promotes investigation related to the full range of questions concerning merit (not just those related to desired outcomes or future decisions). The needed definition should also utilize all potentially relevant methods (not just standardized tests, experimental designs, or panels of experts).

The essential definition that I propose is one that appears in standard dictionaries. It is that:

> Evaluation is the act of examining and judging, concerning the worth, quality, significance, amount, degree, or condition of something.

In short, evaluation is the ascertainment of merit. This definition meets the basic conditions of considering all questions related to worth and is not restricted to the use of any particular methodology. Also, it has currency since it reflects common usage of the term evaluation.

The weakness of this definition is that it provides almost no operational guidance for conducting evaluation. For this reason, I will return to the matter of definition when I respond to my tenth question about evaluation.

Now for the second question:

2) What are the various objects of evaluation?

Evaluation is a pervasive concept that can and should be applied to all aspects of education. Too often evaluation has been restricted to mean assessing student performance, teacher performance, or the

achievements of special projects. While these are appropriate foci for evaluation work, they are by no means the only ones. It is important to have in mind a broad array of objects for evaluation, not only because they should be subjected to scrutiny, but so that a methodology with broad applicability can be devised.

A partial list of appropriate objects is as follows:

Programs -- which are on-going, goal-directed activities, involving especially curricular offerings

Projects -- which are time-bounded, objective-directed activities such as federally funded workshops

Personnel -- which are the people--such as teachers, administrators, and custodians--who perform certain roles

Students -- which are the individuals enrolled in instructional offerings

Courses -- which are the basic modules within which instruction for given groups of students is provided

Facilities -- which are the non-content physical items such as buildings, grounds and equipment

Materials -- which are the content-related physical items, including books, films, and tapes

Institutions -- which are entire organizational entities such as colleges, school districts, community service agencies, and foundations

Budgets -- which are plans for coordinating resources and expenditures

Evaluations -- which are attempts to ascertain the merit of something.

This list illustrates that evaluation is a pervasive enterprise. The remainder of the conceptual framework to be presented here is intended to be sufficiently general to guide assessments of all of the objects identified above.

Next we turn to the third question: What audiences should evaluations serve?

Generally, evaluations serve multiple audiences. School district evaluation reports may be of interest to students, parents, teachers, principals, superintendents, counselors, boards of education, and the public. As a rule, evaluations should provide information to all people who will be affected by and/or involved with the object to be evaluated.

In designing an evaluation study it is important to analyze the data requirements of the various evaluation audiences. This is because the reports from an evaluation study designed to serve one audience, say a superintendent, likely will not answer all the questions of another audience, say a teacher. Evaluators must thus define their audiences early in the designs of evaluations, interact with them to identify their questions, systematically produce information for them, and then help them to use it.

For our fourth question, we consider how evaluation reports should be used.

In my view evaluation reports have two main uses. These are providing information for decision making and for accountability.

The former requires proactive application of evaluation, as information in this case is provided to decision makers before they make their decisions. Hence the criteria for judging evaluation include relevance to the decisions to be served and time limits that reflect when information is needed. In general, this type of evaluation is equivalent to formative evaluation as defined by Michael Scriven.[2] Formative evaluation helps in developing programs and insuring that they will succeed.

But as Lessinger and others have argued, servicing decision making is not evaluation's sole role: providing information for

accountability is also important. This is a retroactive application of evaluation that provides information after work has been completed, and after all of the implementation decisions have been made. This kind of information generally does not help in producing a success, at least not the first time through. But it can aid those who conducted the enterprise to describe and defend their efforts and accomplishments. It can also assist those who paid the bills, e.g., the taxpayers, and those who were directly serviced by the effort, the consumers, to hold the service-agents responsible for the extent and quality of their work. Evaluation for accountability is similar to what Scriven has termed summative evaluation.[3]

Now we turn to our fifth question: What questions are addressed in evaluation work?

As I have argued elsewhere,[4] evaluation studies address four questions. How good are the goals? How good is the plan for achieving the goals? How well has the plan been executed? How good are the outcomes?

These questions take on added meaning as we identify the object of an evaluation. For example if we are evaluating a teacher we might ask the following questions: How explicit and defensible are his year-long and unit objectives? How complete, relevant and feasible are his instructional plans? How adequately is he carrying out his plans? What are the outcomes of his teaching and how do these respond to student needs and to his own objectives?

The four questions take on even further meaning when they are juxtaposed to the two uses of evaluation that we considered a moment ago. When the purpose is to serve decision making the evaluation addresses

specific questions that are of interest to those persons who must formulate objectives, plan how they will be achieved, implement their plans, and decide whether to terminate, modify or continue their effort. Conversely, when the purpose is to serve accountability, the evaluation must address specific questions that are of interest to the clients and sponsors of the programs. They likely would want to know what human needs were addressed, whether accepted practices were employed, whether funds were expended wisely, and whether the outcomes were as good as those from some comparable program. Some examples should further clarify how the four general evaluation questions interact with the two main uses of evaluation.

Suppose that the purpose of an evaluation study is to decide on goals for a new remedial reading program. In this case the evaluation should supply information to assist in formulating appropriate goals. Here the evaluation would address questions such as the following: What do diagnostic tests reveal about the reading <u>needs</u> of the group to be served? What special <u>problems</u> (such as language barriers and health difficulties) have to be solved before the students' reading needs can be met? And, what funding and other kinds of <u>opportunities</u> are available to meet the needs and to solve the problems? Timely responses to these questions would aid decision makers to develop goals for the new reading program that respond opportunistically to the students' needs.

As another example, suppose that the purpose of the evaluation is to serve accountability and that the main question is whether the plan for a new mathematics program has been adequately implemented.

The audience for such an accountability study likely would require answers to the following questions: Was the operational plan carried out? How well did the staff fulfill their responsibilities? Were expenditures made in accordance with the budget? Responses to questions such as these help an audience determine whether a plan has been executed sufficiently well that it is appropriate to judge it by its outcomes.

Time does not permit a fuller explication of the four main questions that are addressed in evaluation studies. Once again, they concern goals, plans, implementation and results.

Next we consider question number six: What information does evaluation require?

According to Robert Stake,[5] evaluators should obtain both descriptive and judgmental information.

In the case of descriptive information, Stake calls for a full portrayal of what is being evaluated. He suggests that an enterprise be described in relation to both its intended and actual characteristics. For example, he emphasizes that results be decribed in regard to those that were intended (the goals) and those that were not anticipated (the side effects). The point to be underscored is that evaluations should be done in relation to questions about an object's goals, plan, execution and results.

In addition to descriptive information, Stake emphasizes that evaluators should collect judgmental information. Stake's position is that persons of all walks who interact with the object of the evaluation, whether it be a teacher or a building, potentially can offer valuable judgments about that object. I concur and suggest that evaluation designs

provide for collecting judgments from clients, the implementors, the governing bodies, appropriate experts, and the evaluators.

So far in this conceptualization of evaluation I have noted 1) that evaluation is the ascertainment of merit, 2) that it can be applied to a variety of objects, 3) that it serves multiple audiences, 4) that it serves both decision making and accountability, 5) that it addresses questions about goals, plans, execution, and results, and 6) that it uses both descriptive and judgmental information.

For the seventh question we consider: Who should do evaluations?

Whether "insiders" or "outsiders" should do evaluations is a question often posed in discussions of evaluation. In general, internal evaluators are those who are directly employed by the agency that houses the teacher, program, project or other object being evaluated. Conversely, external evaluators are, as a rule, employed outside the agency whose work or personnel are being assessed. While this is an oversimplification it is sufficient to introduce three points that I wish to make.

First, an in-house evaluation group should perform the formative, proactive evaluation that is needed to guide decision making. This is because insiders are in a better position than are outsiders to provide the intensive, constructive criticism that is needed to guide programs.

On the other side, outsiders are needed to provide accountability reports. This is because they can be more objective and credible to outside audiences, than can insiders.

Finally, internal, proactive evaluation for decision making is fundamentally more important than external evaluation for accountability.

Let me elaborate briefly on
evaluation can survive better wit
This is because external evaluati
data on previously conducted inte
internal, formative evaluation ca
But, external, summative evaluati
was previously achieved.

Now for the eighth question:

Most evaluators would agree
interactive process with no set s
written on the topic have propose
what tasks must be performed. It
in mind, as they provide an essen
since they must be considered in
studies.

The process that I propose i
delineating, obtaining and applyi
and applying stages involve inter
their audiences, either to collab
or in applying the obtained data
questions. The other stage, obta
activity, and it involves data co
Let me describe each of these sta

Delineating pertains to the
Included in this stage are the fo
identifying the object to be eva

be served, specifying their purposes for the evaluation, identifying the questions to be answered, clarifying the assumptions to be made about design, measurement and analysis, defining the system in which the evaluation is to occur; and, finally, determining the policy and contractual agreements that will govern the evaluation work. Agreements in these areas should be reached in the initial planning of an evaluation study, and these agreements should be open to renegotiation throughout the study.

The second stage in the evaluation process includes all the technical activities that are required to obtain the needed information. These include: specifying the samples for data gathering, choosing appropriate data gathering instruments, specifying the data gathering procedures, determining the means of data storage and retrieval, selecting the analysis techniques, and outlining the required technical reports. These steps require little interaction between evaluators and audiences. But the proper implementation of these design steps can require considerable technical expertise. This is why competent evaluation usually requires the assistance of evaluation specialists.

Another point to be made about the technical side of evaluation is that it involves a wide array of methods. These include such classical techniques as surveys, site visits, systematic observations, checklists, standardized tests, and experiments. Newer evaluation techniques are adversary hearings, advocacy reports, traveling observers, goal-free evaluation, modus operandi analysis, and meta-analysis. These and other techniques provide a rich reservoir that evaluators should use in designing and conducting their studies.

Application is the third and final stage in the evaluation process. Following the obtaining of information, evaluators and their audiences need to collaborate in applying the data to pertinent decision-making and accountability functions. The evaluator's essential role in this stage is to provide information to the appropriate audiences. The role of the audiences is to use the information for decision making and accountability.

Next we consider the ninth question: By what standards should evaluations be judged?

By now it should be obvious that evaluation is a complex and difficult area, and that it may be done well or badly. Also, it can be a threatening and destructive process. Hence, evaluation itself, which I shall term primary evaluation, is an important object of evaluation.

The literature of evaluation refers to the evaluation of primary evaluation as either meta-evaluation or auditing. I will use the former of these labels.

As with primary evaluation--such as I have so far discussed--meta-evaluation has both proactive and retroactive features. Meta-evaluation should both guide on-going evaluation efforts and judge their merits once they have been completed. It should provide both descriptive and judgmental information about the goals, design, implementation, and results of the evaluation study. It should serve the evaluators and their audiences, and it should be done by the primary evaluators themselves and by outsiders who would serve as special meta-evaluators. Finally meta-evaluation involves the same process as does primary evaluation. One way persons whose work is being evaluated can protect themselves

from inept or unscrupulous evaluators is to insist on an impartial third party meta-evaluation of the primary evaluation.

In conducting meta-evaluation it is important to have in mind appropriate criteria for judging evaluation reports. These include the following criteria of technical adequacy: reliability, validity and objectivity. They also include the utility criteria of scope, relevance, importance, credibility, timeliness, and pervasiveness. They include the ethical criterion of respecting peoples' rights, and they include a concern for cost/effectiveness of the evaluation results. Hence, evaluations should be judged according to whether they provide accurate information, whether the evaluation serves useful purposes, whether it is obtained and reported in ethical ways, and whether it is obtained at a reasonable cost. Evaluators who satisfy these conditions are doing their work well.

So far, I have proposed responses to nine questions about evaluation. In summary my responses are (1) evaluation is the ascertainment of merit, (2) it may be applied to a variety of objects, (3) it serves multiple audiences, (4) it serves both decision making and accountability, (5) it assesses goals, plans, activities, and results, (6) it provides both descriptive and judgmental information, (7) it is conducted by both insiders and outsiders, (8) it follows a process of delineating, obtaining, and applying the evaluative information, and (9) it should itself be evaluated for its technical adequacy, utility, ethical treatment of people, and cost/effectiveness.

Given these nine responses, I now consider the tenth and final question: What operational definition should guide evaluation work?

At the outset of this discussion of ten questions about evaluation, I noted that evaluation is the ascertainment of merit. This definition was defended for its breadth in covering all important questions of merit and its flexibility in considering all relevant methods. Its main limitation is its generality. Hence at this point it is appropriate to attempt an operational definition that is consistent with the general one and with the theoretical framework thus far developed in this paper.

My proposed operational definition is as follows:

> Evaluation is the process of delineating, obtaining and applying descriptive and judgmental information; concerning some object's merit; as revealed by its goals, plans, process and product; and for the purposes of decision making and accountability.

This definition conforms to the essential dictionary meaning of evaluation, since it focuses on the assessment of merit. And, it provides guidance for organizing and conducting evaluation studies.

Concluding Remarks

In this paper I have described a conceptual framework that educational administrators may want to adopt as a general guide for their evaluative activities. While it is not crucial that educational administrators adopt all of the positions argued in this paper, it is essential that they insure and demonstrate that their programs have real merit. I believe that these purposes are best served through a sound evaluation program. I hope that some of my suggestions may aid administrators to design and implement such evaluation programs.

References

[1]Stufflebeam, Daniel L. Meta-evaluation. The Evaluation Center Occasional Paper Series, 3, Western Michigan University, May, 1975.

[2]Scriven, Michael S. "The Methodology of Evaluation." Perspectives of Curriculum Evaluation (AERA Monograph Series on Curriculum Evaluation, No. 1). Chicago: Rand McNally & Co., 1967.

[3]Ibid.

[4]Stufflebeam, Daniel L. "The Relevance of the CIPP Evaluation Model for Educational Accountability." SRIS Quarterly 5. Spring, 1972.

[5]Stake, Robert E. "The Countenance of Educational Evaluation." Teachers College Record, Vol. 68, 1967.